STAT 651 Chapter 1

Zepu Zhang
December 16, 2010

1 Basic set theory

Basic concepts

Set A.

Object x.

Relation between x and A:

$$\text{either } x \in A \text{ or } x \notin A$$

"membership"—x is (not) a member/element of A; x is (not) in A; x belongs to (does not belong to) A.

Whole set S, say.

Empty set $\emptyset$.

Subset: $A \subset S$.

Complement: A^c.

Example 1. Cards: each card, spades, hearts, diamonds, suits, A's, 2's,...
2. Class: each student, male, female, undergrad, grad,...

Set operations

Operations on sets:
$A \cup B$
$A \cap B$
A^c

Relations between sets:

(1) subset, or "containment": $A \subset B$ ("A is contained in B" or "A is a subset of B")

(2) equality: $A = B$, defined as $A \subset B$ and $B \subset A$

(3) disjoint, or mutually exclusive: $A \cap B = \emptyset$

(4) overlapping: $A \cap B \neq \emptyset$ ("Overlap" is not a standard term, but this certainly is an important relation.)

(5) partition: $A \cap B = \emptyset$ and $A \cup B = S$

Exercise $A = \big\{1, \{2,3\}, \{4,5,6\}, \{7,8\}\big\}$; $B = \{4,5,6\}$. What's the relation between A and B? $B \in A$ or $B \subset A$? What if $B = \{1\}$? And what if $B = \big\{\{1\}\big\}$?

Theorem 1.1.4. **Rules for set operations**:
(1) commutativity
(2) associativity
(3) distributive laws
(4) DeMorgan's laws

Important things to learn

All content of Chapter 1.1 is required. In particular note the following technical points.

1. Comfortably use the symbols $\in$, $\subset$, $\cup$, $\cap$, ^{c} to denote set relations and operations.

2. Use rigorous math symbols to denote sets. For example, $A \cup B$ is defined as the set

$$\{x : x \in A \text{ or } x \in B\},$$

whereas A^c is defined as the set

$$\{x : x \notin A\},$$

where $x \in \Omega$ (the whole set) is implied.

3. Explicitly define a set by enclosing its members between a pair of braces. Two situations:

   (a) Explicitly list all members, e.g. $\{1,3,5,7\}$, $\{\text{Lisa}, \text{Mary}, \text{Jane}\}$, $\{1,2,3,...\}$.

   (b) Give a description of the members, e.g. $\{x : \sqrt{x} > 3 \text{ and } x^2 < 100\}$. (It is usually implicit that x is a real number.)

4. Use definition to prove set relations, esp set equality.

5. Use Venn diagrams to help see, understand, discover, and prove set relations.

Example Prove the distributive law $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. See page 3.

Example Prove the associativity property $A \cup (B \cup C) = (A \cup B) \cup C$.

Proof

First, prove $A \cup (B \cup C) \subset (A \cup B) \cup C$. This is equivalent to saying that if $x \in \big(A \cup (B \cup C)\big)$, then $x \in \big((A \cup B) \cup C\big)$.

Since $x \in \big(A \cup (B \cup C)\big)$, then $x \in A$ or $x \in (B \cup C)$.
(1) If $x \in A$, then $x \in (A \cup B)$, hence $x \in \big((A \cup B) \cup C\big)$.
(2) If $x \in (B \cup C)$, then $x \in B$ or $x \in C$.
(2a) If $x \in B$, then $x \in (A \cup B)$, hence $x \in \big((A \cup B) \cup C\big)$.
(2b) If $x \in C$, then $x \in \big((A \cup B) \cup C\big)$.

Second, prove $(A \cup B) \cup C \subset A \cup (B \cup C)$. This is equivalent to saying that if $x \in \big((A \cup B) \cup C\big)$, then $x \in \big(A \cup (B \cup C)\big)$.

Since $x \in \big((A \cup B) \cup C\big)$, then $x \in (A \cup B)$ or $x \in C$.
(1) If $x \in (A \cup B)$, then $x \in A$ or $x \in B$.
(1a) If $x \in A$, then $x \in \big(A \cup (B \cup C)\big)$.
(1b) If $x \in B$, then $x \in (B \cup C)$, hence $x \in \big(A \cup (B \cup C)\big)$.
(2) If $x \in C$, then $x \in (B \cup C)$, hence $x \in \big(A \cup (B \cup C)\big)$.

The proof is now complete.

Extensions

Extend the $\cup$, $\cap$ operations and the mutually exclusiveness and partition concepts to multiple (finite or infinite) sets.

Finite, infinite, countable, uncountable.

Example (1) $\cup_{i=1}^{\infty} \big[(1/i), 1\big] = (0, 1]$
(2) The sets $[i, i+1)$, $i = 0, 1, \ldots$, form a partition of $[0, \infty)$.

Connections between set theory and probability theory

Experiment and outcomes.

Experiment: an action or procedure whose outcome is uncertain.

When the outcome contains multiple components, it is important to know/specify whether their order matters.

Definition 1.1.1. **Sample space**, $\mathcal{S}$, of an experiment: the set of all possible outcomes of the experiment.

Exercise List the sample space of
(1) toss a coin once;
(2) toss a coin 3 times;
(3) toss 3 coins simultaneously;
(4) throw a die twice;
(5) draw 2 cards and get their types;
(6) take a measure of the temperature (the randomness is due to measurement; the temperature at any given moment

is assumed to be an unknown constant).

Definition 1.1.2. **Event**: any collection (i.e. set) of outcomes in the sample space.

Note: an event is a set; its members are "outcomes".

An event occurs: if the outcome is in the collection of outcomes that define the event. Therefore, one experiment may trigger multiple events to occur.

Exercise Define some events in the previous exercise.

2 Basics of probability theory

2.1 Axiomatic foundations

In STAT 200 and STAT 300, we understood "probability" as "relative frequency of occurrence in repeated experiments".

Now take a different, more general, somewhat more abstract, and more defensible approach. In essence, we define what "probability" should look like (and declare that anything with that look is valid probability), but do not stipulate what "probability" means. This is the "axiomatic approach".

Definition 1.2.1. **Sigma algebra** (Borel field).

From the remarks below definition 1.2.1 on page 6, appreciate an important requirement on mathematical definitions—a definition should eliminate redundancy. A typical way to define something is to list a necessary and sufficient group of properties of it. We sometimes see different ways to define the same thing, because different authors prefer different groups of defining properties. But the definitions are all equivalent.

Several basic facts derived from the definition:

1. $S \in \mathcal{B}$.
2. $\mathcal{B}$ is closed under countable intersections.
3. $\mathcal{B}$ is closed under finite unions and intersections.

Example 1.2.2

Example 1.2.3

Definition 1.2.4. **Probability function**.

The "domain" $\mathcal{B}$ of the function is a sigma algebra, which is a collection of "events". The "range" of the function is $[0, 1]$. The probability function takes any element of $\mathcal{B}$

(that is, any event) as input, and gives a single real number on $[0,1]$, called the "probability" of that event.

Requirement 3 has a name: "countable additivity".

This definition chooses three properties as defining properties of the concept. Other properties can be derived from these three. One may ask, upon seeing the definition, (1) What is $P(\emptyset)$? (2) Is it true that $P(A) \leq 1$ for any set A? (3) If we know $P(A)$, can we calculate $P(A^c)$? These are answered in **theorem 1.2.8**.

Example 1.2.5

The definition of probability function does not specify the meaning of probability. There are two interpretations to a probability: (1) relative frequency of occurrence of the event in repeated experiments; (2) some sort of "subjective belief" in this event on the scale $[0,1]$. (Naturally, 0 means impossible and 1 means definite.)

So the definition does not specify (or restrict) where the probability function comes from, then how can we define a probability function? By defining a probability function, we mean assigning a probability (a single number on $[0,1]$) to each event (that is, defining a function from $\mathcal{B}$ to $[0,1]$) such that the function satisfies the three axioms.

Any such function is a valid probability function, according to the definition. We define it in a way that is "natural" and "sensible". Here we follow the "frequency" path only.

The recipe is based on

Theorem 1.2.6

and the following: if two outcomes s_i and s_j are "equally likely", then the two events $\{s_i\}$ and $\{s_j\}$ have the same probability.

Consequently, if $S = \{s_1, \ldots, s_n\}$ is finite and the outcomes $s_1, \ldots, s_n$ are all "equally likely", then $P(\{x_i\}) = 1/n$ for $i = 1, \ldots, n$. See **example 1.2.5**.

If an experiment have N possible outcomes and all the outcomes are <u>equally likely</u>, then $P(A) = N(A)/N$, where $N(A)$ is the number of outcomes in event A.

Example Draw one card from a fully mixed deck of cards. There are 52 possible outcomes and it is reasonable to assume all individual cards have the same chance of being picked. Event A: the card drawn is a spade.

Answer: $N(A) = 13$, hence $P(A) = 13/52 = 1/4$.

If the outcomes are "continuous", the idea is analogous. See **example 1.2.7**.

Question: in example 1.2.7, what is the sample space?

2.2 The calculus of probabilities

Theorem 1.2.8

Theorem 1.2.9

Theorem 1.2.11

Prove these theorems. Illustrate with Venn diagrams (as always).

These are all very useful, for example in the subsequent sections on "counting". For example, if $P(A)$ is hard to find, find $P(A^c)$ and use 1.2.8c. For another example, find $P(B)$ using 1.2.9a, or $P(A)$ using 1.2.11a.

1.2.11a is an extension/generalization of 1.2.9a. It is used in classifying the question into smaller, disjoint ones, and performing divide-and-conquer.

1.2.11b is somewhat a generalization of 1.2.9b.

Perhaps 1.2.9a is more naturally written as
$P(A \cap B) = P(B) - P(B \cap A^c) = P(A) - P(A \cap B^c)$
or $P(B) = P(B \cap A) + P(B \cap A^c)$.

1.2.9b will be used a lot.

2.3 Counting

Theorem 1.2.14. **Fundamental theorem of counting**, or product (or multiplication) rule.

With replacement; without replacement.

Ordered; un-ordered.

Example Throw 4 dice into 4 spots. How many possible outcomes?

Example Draw 3 cards from one deck and arrange them in the order they are drawn. How many possible outcomes?

Example Experiment: draw 4 cards.
Event: all 4 cards are of the same suit, and their ranks (face number) are in sequence, like 2, 3, 4, 5.
Question: how many possible outcomes does this event con-

tain?

Answer

Think of it as 2 steps: (1) choose the suit—4 choices; (2) choose the ranks—11 choices (1,2,3,4; 2,3,4,5;...;11,12,13,1). $4 \times 11 = 44$.

Or think of it this way: each outcome is uniquely identified by an ordered pair of items: `(suit, starting rank)`.

Definition 1.2.16. **Factorial**:

$$n! = \cdots$$

There are $n!$ possible orderings for n items.

Definition **Permutation**: choosing r out of n objects without replacement, and the order matters. How many possible outcomes?

$$P_r^n = n \cdot (n-1) \cdot (n-2) \cdots [n-(r-1)] = \frac{n!}{(n-r)!}$$

Product of r numbers counting down from n. (There are different ways to define the P notation.)

Definition 1.2.17. **Combination**: choosing r out of n objects without replacement, and the order does not matter. How many possible outcomes (that is, unique combinations)?

$$\binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}$$

(To turn a permutation into a combination, ignore the difference between the $r!$ orderings.)

Example Page 14–15, after definition 1.2.16.

Understand item 4 on page 15:
(1) Why $44^6/6!$ is wrong, and is too small? See **example 1.2.19**.
(2) Understand the formula $\binom{n+r-1}{r}$: think of $n+r-1$ objects ($n-1$ walls plus r markers); pick r of them as markers.

Table 1.2.1. (The "ordered, without replacement" case is permutation.)

2.4 Enumerating outcomes

(For Section 1.2.4, only read through the first paragraph of page 18.)

In an experiment with equally likely outcomes, calculating probabilities amounts to counting the number of outcomes in the desired event, plus counting the number of all possible outcomes. (This is what the title of this section means.)

Example A quality control engineer is to randomly pick 5 out of the 11 available products for inspection. Suppose 2 of the 11 products have defects. What is the probability that (1) no defective product is picked? (2) exactly one defective product is picked?

Example 1.2.18

Example 1.2.19

Example Consider a regular deck of 52 playing cards. For a five-card poker hand, find the probability of (a) All of different ranks; (b) One pair; (c) Two pairs; (d) Three of a kind: three cards of the same rank and two others of different ranks, for example JJJ74; (e) A straight: five cards in sequence; the ace can be either high or low; (f) A flush: five cards of the same suit; (g) All are spades; (h) There are exactly 2 clubs and 2 hearts.

Example A box in a certain supply room contains four 40-W lightbulbs, five 60-W bulbs, and six 75-W bulbs. Suppose that three bulbs are randomly selected. (*) What is the probability that exactly two of the selected bulbs are rated 75 W? (**) What is the probability that all three of the selected bulbs have the same rating? (***) What is the probability that one bulb of each type is selected?

Answer:

(1) $\binom{9}{5} \Big/ \binom{11}{5}$. (2) $\binom{2}{1}\binom{9}{4} \Big/ \binom{11}{5}$.

(a) $\binom{13}{5}4^5 \Big/ \binom{52}{5}$. (b) $13\binom{4}{4}\binom{12}{3}4^3 \Big/ \binom{52}{5}$.

(c) $\binom{13}{2}\binom{4}{2}^2\binom{44}{1} \Big/ \binom{52}{5}$.

(d) $13\binom{4}{3}\binom{12}{2}4^2 \Big/ \binom{52}{5}$. (e) $11 \times 4^5 \Big/ \binom{52}{5}$.

(f) $4\binom{13}{5} \Big/ \binom{52}{5}$. (g) $\binom{13}{5} \Big/ \binom{52}{5}$.

(h) $\binom{13}{2}^2 \Big/ \binom{52}{5}$.

(*) $\binom{6}{2}\binom{9}{1} \Big/ \binom{15}{3}$.

(**) $\left[\binom{4}{3}+\binom{5}{3}+\binom{6}{3}\right] \Big/ \binom{15}{3}$. (***) $4\cdot5\cdot6 \Big/ \binom{15}{3}$.

3 Conditional probability and independence

Example Roll a die, $A = \{1\}$. $B = \{1, 3, 5\}$. (1) $P(A)$? (2) $P(B)$? (3) $P(A \mid B)$?

Answer: $P(A) = 1/6$. $P(B) = 1/2$. $P(A \mid B) = 1/3$.

Why?

In the case of equally-likely outcomes,

$$P(A \mid B) = \frac{N(\text{in } A\text{, besides being in } B)}{N(\text{in } B)} = \frac{N(A \cap B)}{N(B)}$$

Since B is a condition, that is, it is assumed that B occurs, then the sample space for the subsequent (conditional) event is B. Under this condition, the event A is actually $A \cap B$. Hence $P(A \mid B) = N(A \cap B)/N(B)$.

Definition 1.3.2. **Conditional probability**:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Of course,

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

is the same thing, and is also useful.

Use Venn diagram to illustrate the concept.

The rearrangement,

$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$$

if very useful. Interpret it in words.

Example 1.3.1

Example 1.3.3. First, figure out the final answer directly. Second, use conditional probability.

Example Draw two cards from a deck of 52 cards. What's the probability of getting a pair?

Answer

Solution 1. $N(S) = \binom{52}{2}$. $N(A) = \binom{13}{1}\binom{4}{2}$.

$$P(A) = \frac{N(A)}{N(S)} = \frac{13 \times 4 \times 3}{1 \times 2} \bigg/ \frac{52 \times 51}{1 \times 2} = 1/17$$

Solution 2.

$$A = (1,1) \cup (2,2) \cup \cdots \cup (13,13)$$

These sub-events are mutually exclusive, and they all have the same probability.

$$\begin{aligned}
P(A) &= \sum_{i=1}^{13} P(i,i) \\
&= \sum_{i=1}^{13} P(\text{first draw } i)P(\text{second draw } i \,|\, \text{first draw } i) \\
&= 13 \times \frac{4}{52}\frac{3}{51} \\
&= \frac{1}{17}
\end{aligned}$$

Example Pick 2 cards without replacement from a 52 deck.
A1 = an ace is selected on the first draw.
A2 = an ace is selected on the second draw.
Find $P(A_1)$, $P(A_2)$.

Answer

$$\begin{aligned}
P(A_1) &= \frac{4}{52} \\
P(A_2) &= P(A_1)P(A_2 \,|\, A_1) + P(A_1^c)P(A_2 \,|\, A_1^c) \\
&= \frac{4}{52}\frac{3}{51} + \frac{48}{52}\frac{4}{51} = \frac{4}{52}
\end{aligned}$$

We used the relation $P(A_2) = P(A_1 \cap A_2) + P(A_1^c \cap A_2)$. This is a very general and powerful "classify and tackle them one by one" strategy. Note: $A_2 = A_1$!
Implication: in a poker game, seating should not matter.

Example The probability a randomly selected family belongs to the AAA auto club is 0.25 (*Source*: American Automobile Association). If a family belongs to AAA, the probability they have more than one car is 0.45. Suppose a family is randomly selected. What is the probability they have more than one car and belong to AAA?

Answer

A = belong to AAA. M = have more than one car.

$$P(M \cap A) = P(A)P(M \,|\, A) = 0.25 \times 0.45 = 0.1125.$$

It is also true that $P(M \cap A) = P(M)P(A \,|\, M)$, but it is not useful for this question, because $P(M)$ and $P(A \,|\, M)$ are unknown.

Example During frequent trips to a certain city a traveling salesperson stays at hotel A 50% of the time, at hotel B 30% of the time, and at hotel C 20% of the time. When checking in, there is some problems with the reservation 3% of the time at hotel A, 6% of the time at hotel B, and 10% of the time at hotel C. Suppose the salesperson travels to this city.
(a) Find the probability the salesperson stays at hotel A and has a problem with the reservation.
(b) Find the probability the salesperson has a problem with the reservation.
(c) Suppose the salesperson has a problem with the reservation, what is the probability the salesperson is staying at hotel A?

In the formula of definition 1.3.2, it is often easier to find $P(A \cap B)$ via the relation $P(A \cap B) = P(A)\, P(B \,|\, A)$, which is just a re-use of definition 1.3.2. Therefore,

$$P(A \,|\, B) = \frac{P(A)\, P(B \,|\, A)}{P(B)}$$

This is called the **Bayes' rule**.

This rule deals with this situation: Originally we have some knowledge about A (represented by its probability), without info about B. Now we are told B has happend. Given this info, is our knowledge about A improved? Intuitively, yes, if A and B affect each other. Then, how do we update our knowledge of A, given the ocurrence of B? Well, our updated knowledge is $P(A \,|\, B)$, and the updating is obtained by Bayes' rule.

In Bayesian nomenclature, $P(A)$ is the prior probability of A. The prior is updated to the posterior, $P(A \,|\, B)$, in light of the likelihood $P(B \,|\, A)$.

The rule can also be obtained by re-arranging

$$P(A)\, P(B \,|\, A) = P(B)\, P(A \,|\, B).$$

An extended form of the Bayes' rule is given in **theorem 1.3.5**. The denominator in theorem 1.3.5 is equal to $P(B)$, according to theorem 1.2.11a.

How can I memorize the Bayes' rule?

Example 1.3.6.

If $P(A \,|\, B) = P(A)$, that is, the probability of A's occurrence is not affected by whether B occurs, we say A and B are independent.

Definition 1.3.7. **Independence**.

Note:

1. Note the three equivalent forms: (1) $P(A \mid B) = P(A)$; (2) $P(B \mid A) = P(B)$; (3) $P(A \cap B) = P(A)\,P(B)$. Verify that the three forms are equivalent.

2. The definitions are also properties. If we know A and B are independent, then we know the three relations above hold. The third is used especially often.

3. Independence is not the same as disjoint. Actually if $A \cap B = \emptyset$, the two events are dependent. Knowing that B has occurred certainly tells us about the occurrence of A: A will not occur.

4. There is no useful way to represent "independence" by a Venn diagram.

5. The third definition is generalized to a definition of independence between more than two events. See **definition 1.3.12**.

6. In the definition of independence, we can replace any event by its complement. For example, if A and B are independent, then A^c and B are independent. See **theorem 1.3.9**.

Example 1.3.8.

4 Random variables

We've been talking about experiments, outcomes, events (i.e. sets of outcomes) etc. There are all kinds of experiment-outcomes, for example:

type of outcome	example
number	throw a die once, get the face number
character	pick a student randomly, get the name
a group of things	the first 3 people who enter Chapman bldg on Friday
an ordered list	at University and College, the colors of the first 10 vehicles after noon
a description	low/normal/high blood pressure of a random patient

We want to unify things under the roof of "numbers"; specifically, real values. After that we will build a whole bunch of tools to characterise and study them.

Definition 1.4.1. **Random variable**: a function from a sample space **into** the real numbers.

Note "into"—the range of the function does not have to be the whole $(-\infty, \infty)$. It may be just a few numbers; or all integers; or certain interval(s); or $(-\infty, \infty)$, or some subset of $(-\infty, \infty)$, and so on.

There are many possibilities of the definition of this function, for example:

types of outcome and rv X	example
a number; directly	throw a die; face number
	measurement of temperature; measured temperature
complicated; a summary	the list of H/T of 10 coin tosses; number of heads
	survey of 1000 people; number of 'Yes'
	applicant's credentials; ACT score
non-numerical; coding	success vs failure; 1 vs 0

The essence of the concept of function is that the input uniquely determines the output. Did the above define functions?

Extension: the concept of multivariate random variable, or random vector.

The rv has its own **sample space**, i.e. the range of the function, and a probability function defined on a sigma algebra of its sample space.

S: sample space of the experiment.
s: any outcome, $s \in S$.
$X(s) : S \to \mathcal{R}$.
$\mathcal{X}$: sample space (or range) of X.

$$\mathcal{X} = \{x : X(s) = x, s \in S\}$$

A probability function, $P_X(x)$ where $x \in \mathcal{X}$, is induced from that of the original experiment:

$$P_X(X = x) = P\big(\{s \in S : X(s) = x\}\big)$$

Example 1.4.3.

Example 1.4.4.

5 Distribution functions (cdf)

Definition 1.5.1. $F_X(x) = P_X(X \le x)$, for all x.

Attention: $\le$. right continuity.

Example 1.5.2. (Esp. note how the plot indicates right continuity and that the function is defined on the whole real line.)

Observations below example 1.5.2:
1. F_X is defined for all $x \in (-\infty, \infty)$, even if S_X does not contain all real numbers.
2. F_X jumps at the values $x_i \in \mathcal{X}$, size of the jump being $P(X = x_i)$.
3. $F_X(x) = 0$ for $x < \min(x)$, $F_X(x) = 1$ for $x > \max(x)$.
4. $F_X(x)$ is flat between adjacent x values if X is discrete.

Exercise How do you formally define or describe "right continuity"? How would you go about proving that a cdf is right continuous?

Theorem 1.5.3. $F(x)$ is a cdf iif
(1) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
(2) nondecreasing, i.e. $x_1 \geq x_2 \Rightarrow F(x_1) \geq F(x_2)$.
(3) right continuous, i.e. $\lim_{x \downarrow x_0} F(x) = F(x_0)$.

Example 1.5.4. Discrete distribution.

Example 1.5.5. Continuous distributions.

Example 1.5.6. cdf with jumps. (Illustrate with a graph.)

You should be familiar with the look of cdf curves, and know what these features mean for the distribution:
(1) step function, right continuous;
(2) continuously increasing;
(3) with jumps;
(4) with flat segments.
(5) mix of continuously increasing, jumps, flats.

Definition 1.5.7. Continuous and discrete rv's.

Observations on the definition:

> X is continuous if $F_X(x)$ is continuous—shouldn't have jumps, that is, no value of X has a nonzero probability (point mass). But flat segments are allowed, meaning the range of X could consist of disjoint intervals.
>
> X is discrete if $F_X(x)$ is a step function—no continuous segments.
>
> According to this definition, a mixture of continuous and discrete distributions is neither continuous nor discrete. Although easily conceivable, such mixture distributions is not studied in this course.

Important: **the cdf $F_X(x)$ completely determines the probability distribution of a rv X.**

Definition 1.5.8. What do we mean by saying X and Y have the same distribution?

Theorem 1.5.10. How do we tell whether X and Y have the same distribution?

6 Density and mass functions (pdf and pmf)

Definition 1.6.1. pmf of a discrete rv.

Note

1. pmf gives "point probabilities" or "point masses". Of course, point probabilities can be easily derived from the cdf. pmf and cdf are equivalent—both completely describe the distribution.

$$\text{cdf} \xleftarrow{\text{summation}} \xrightarrow[\text{difference}]{} \text{pmf}$$

2. From pmf to cdf:

$$F(x) = P(X \le x) = \sum_{x_i \le x} P(X = x_i) = \sum_{x_i \le x} f(x_i)$$

3. From cdf to pmf:

$$f(x) = \begin{cases} F(x) - \lim_{t \to x-} F(t), & x \in \mathcal{X} \\ 0, & x \notin \mathcal{X} \end{cases}$$

where $F(x) - \lim_{t \to x-} F(t)$ is the mathematical way to say the "jump" at x. $\lim_{t \to x-}$ is also written as $\lim_{t \uparrow x}$.

Example 1.6.2.

Definition 1.6.3. pdf of a continuous rv.

Note

1. pdf helps give "interval probabilities".

$$\text{cdf} \xleftarrow{\text{integration}} \xrightarrow[\text{differentiation}]{} \text{pdf}$$

2. From pdf to cdf and interval probabilities (based directly on the definition 1.6.3):

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

$$P(a < X \le b) = F(b) - F(a) = \int_{a}^{b} f(x)\,\mathrm{d}x$$

3. $f(x)$ does not tell us the "probability" of X taking the value x! In fact, $f(x)$ can be > 1. For continuous rv X,

   $$P(X = x) = 0$$

   for any x. I think a good way to understand this statement is by observing $P(X = x) = F(x) - \lim_{t \to x-} F(t)$, where the limit is equal to $F(x)$ if $F(t)$ is continuous at $t = x$.

   Consequently,

   $$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$$

4. The definition of pdf does not prohibit "spikes", therefore for any cdf there can be (as far as definition 1.6.3 is concerned) multiple pdf's that satisfy the definition, because spikes do not contribute to the integral. In practice we take the non-spiked, most natural curve—the continuous function.

5. From cdf to pdf:

   $$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

   if $F(x)$ is differentiable and we take the continuous function $f(x)$ among all functions that satisfy the definition 1.6.3.

Example 1.6.4.

Note on notation: examples
random variable X, a specific value x,
cdf $F(x)$, pdf $f(x)$,
distributed as $X \sim F(x)$ or $X \sim f(x)$
identically distributed: $X \sim Y$.

Theorem 1.6.5. Requirements for pdf or pmf.